REMARKS

This Amendment is responsive to the non-final Office Action mailed on November 6, 2003.

I. Objections to the Specification, Claims and Drawings

In the Office Action, the Examiner objected to the use of the term "voice recognition" in the specification and claims, and suggested replacing "voice recognition" with "speech recognition." Because the Examiner also checked the box indicating that the drawings are objected to, Applicant assumes the Examiner objected to the drawings for the same reason.

Because the phrase "voice recognition" appears frequently within the specification, Applicant is making the requested change by submitting a substitute specification—in both clean and redline form pursuant to rules 121 and 125. In addition to replacing "voice recognition" with "speech recognition," the substitute specification replaces "AVR," which stands for Automated Voice Recognition, with "ASR," which stands for Automated Speech Recognition. The substitute specification includes no new matter.

Applicant has also changed "voice recognition" to "speech recognition" in the claims. This change in terminology does not affect the scope of the claims.

Applicant has also submitted replacement drawing sheets for Figures 1 and 3. Figure 1 has been revised solely by changing "voice recognition" to "speech recognition" in block 114. Figure 3 has been revised solely by changing "AVR" to "ASR."

II. Art-based Rejections

The Examiner rejected all of the claims on obviousness grounds over various combinations of the following references: Bennett et al (U.S. Patent 6,615,172), Akers et al (U.S. Patent 6,278,967), Turtle (U.S. Patent 5,265,065), and Malsheen et al (U.S. Patent 5,634,084), collectively referred to herein as "the applied references." For the reasons set forth below, Applicant respectfully submits that the rejection is improper. Applicant will treat Bennett et al and Akers et al as prior art for purposes of responding to the Office Action, but reserves the right to later disqualify one or both references as prior art.

Applicant wishes to initially point out that none of the applied references discloses a method for generating a speech recognition grammar that specifies valid utterances. Although Bennett et al discloses the use of speech recognition grammars, including context-sensitive

grammars, the reference does not describe how these grammars are generated. The remaining three references do not even appear to disclose the use of a speech recognition grammar. (Note that Akers et al appears to use the term "grammar" to refer to linguistic rules and structural relationships for forming sentences, and not to a speech recognition grammar.)

Each independent claim is discussed below:

Independent Claim 1

Independent Claim 1 stands rejected as being unpatentable over Bennett et al in view of Turtle, and in further view of Akers et al. In rejecting Claim 1, the Examiner took the position that Bennett et al discloses a process of incorporating utterances into a voice (speech) recognition grammar, citing column 27, lines 33-36. Applicant respectfully disagrees. The cited portion of Bennett et al merely describes the general function of a speech recognition grammar within Bennett et al's system. As mentioned above, nothing in Bennett et al describes how the grammars are actually generated. Because none of the applied references discloses or suggests incorporating utterances into a speech recognition grammar as claimed, the rejection is improper. See MPEP § 2143.03 (in order to establish prima facie obviousness of a claimed invention, all of the claim limitations must be taught or suggested by the prior art).

Applicant also respectfully submits that the Examiner has failed to identify a valid motivation or suggestion to combine Bennett et al with Turtle and Akers et al. When a rejection depends on a combination of prior art references, there must be some teaching, suggestion, or motivation to combine the references." See, e.g., In re Rouffet, 149 F.3d 1350, 1355, 47 USPQ2d 1453, 1456 (Fed. Cir. 1998). Although a reference need not expressly teach that the disclosure contained therein should be combined with another, the showing of combinability, in whatever form, must nevertheless be "clear and particular." In re Dembiczak, 175 F.3d 994, 999, 50 USPQ2d 1614, 1617 (Fed. Cir. 1999). It is impermissible to "pick and choose from any one reference only so much of it as will support a given position to the exclusion of other parts necessary to the full appreciation of what such reference fairly suggests to one skilled in the art." Bausch & Lomb v. Barnes-Hind/Hydrocurve, 230 USPQ 416, 419 (Fed. Cir. 1986).

In the present case, the Examiner asserts that it would have been obvious to combine Bennett et al with Turtle and Akers et al because all three patents "are from the same field of endeavors, namely speech recognition grammar construction." Office Action at page 8, lines 8-

13. Applicants respectfully disagree. As mentioned above, none of the three references discloses a method of speech recognition grammar construction, and only one of the references even discloses the use of a speech recognition grammar. The Examiner therefore has not met his burden of identifying a motivation to combine.

Applicant further submits that no motivation exists within the applied references themselves to combine the references as proposed by the Examiner. Specifically, nothing in the applied references suggests incorporating either the query parsing operation disclosed in the cited portion of Turtle (namely col. 8, lines 41 and 42), or the sentence subdivision operation described in the cited portion of Akers et al (namely col. 6, lines 56-62), into a process for generating speech recognition grammars. The query parsing operation disclosed in Turtle is performed in order to remove stopwords such as "is," "the" and "of" from a textual, natural language query. The sentence subdivision operation described in Akers et al is performed in order to break a textual sentence into substrings for purposes of classifying the substrings and translating the sentence into another language. Given the very different contexts in which these two operations are described, one skilled in the art would not have been motivated to use these operations to construct speech recognition grammars.

In summary, because the applied references do not disclose all of the limitations of Claim 1, and because no motivation exists to combine the applied references, the rejection of Claim 1 is improper.

Claim 9

The Examiner rejected independent Claim 9 as being obvious over Bennett et al in view of Akers et al. In rejecting Claim 9, the Examiner took the position that Bennett et al discloses "storing at least some of the utterances of the set... within a voice recognition grammar used to interpret the voice-based search query." Office Action at page 3, section 3, citing column 27, lines 33-36 of Bennett et al. Applicant respectfully disagrees. As explained above, the portion of Bennett et al relied on by the Examiner merely describes the general function of a speech recognition grammar, without describing how the speech recognition grammar is generated.

The Examiner also took the position that Akers et al discloses the limitation "translating the phrase into a set of utterances consisting of (a) individual terms of the phrase, and (b) all ordered combinations of two or more consecutive terms of the phrase." Office Action at top of

page 4, citing col. 6, lines 56-62 of Akers et al. Applicant respectfully disagrees. In the example given in Akers et al, the phrase "the man is happy" is not broken into "all ordered combinations of two or more consecutive terms of the phrase" as claimed, as the combinations "man is," "is happy," and "the man is" are omitted.

Because Bennett et al and Akers et al do not disclose or suggest all of the limitations of Claim 9, the rejection is improper.

Applicants also respectfully submit that the Examiner has not identified a valid motivation to combine the teachings of Bennett et al and Akers et al, and that no motivation exists within the references themselves. As explained above, Akers et al does not disclose a method for constructing a speech recognition grammar. The Examiner's assertion that Bennett et al and Akers et al are "from the same field of endeavors, namely speech recognition grammar construction," is therefore inaccurate. In addition, the problem to which the cited portion of Akers et al is directed—namely classifying sentence substrings for purposes of translation to another language—is very different from the problem of selecting utterances to include in a speech recognition grammar. One skilled in the art thus would not have been motivated to incorporate the cited teaching of Akers et al into a process for generating a speech recognition grammar.

In summary, because Bennett et al and Akers et al do not disclose or suggest all of the limitations of Claim 9, and because no motivation exists to combine the two references as proposed by the Examiner, the rejection of Claim 9 is improper.

Claim 16

The Examiner rejected independent Claim 16 as being obvious over Bennett et al in view of Akers et al. Applicant respectfully submits that the rejection is improper, as Bennett et al and Akers et al do not collectively suggest "a grammar which specifies to the speech recognition system valid utterances for interpreting the voice search queries, wherein the grammar comprises both single-term and multi-term utterances derived from the items within the domain, and said multi-term utterances consist primarily of forward combinations derived from phrases within text of the items," within the context of the other claim limitations.

As explained above, Akers et al's teaching of breaking sentences into substrings is disclosed in the context of translating a sentence into another language, and not in the context of

generating a grammar for speech recognition. One skilled in the art therefore would not have been motivated to use the sentence subdivision operation of Akers et al to construct a grammar having the characteristics set forth in Claim 16. The rejection of Claim 16 is therefore improper.

Dependent Claims

The dependent are patentable because of their respective dependencies from the independent Claims 1, 9 and 16. In addition, the dependent claims include limitations that provide additional distinctions over the applied references. By way of example and not limitation, Claims 2, 10 and 17 include limitations involving the use of item titles to generate the grammars. Applicant respectfully disagrees with the Examiner's assertion that these limitations are disclosed at col. 35, lines 33-35 and col. 36, lines 13-16 of Bennett et al.

III. Conclusion

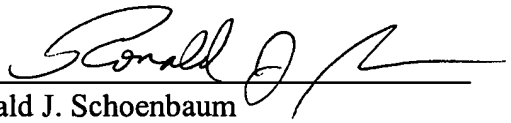
In view of the foregoing remarks, Applicant submits that the bases for objection have been overcome, and that the claims are patentably distinct from the applied references.

If any issues remain which can potentially be resolved by telephone, the Examiner is invited to call the undersigned attorney of record at his direct dial number of 949-721-2950.

Respectfully submitted,

KNOBBE, MARTENS, OLSON & BEAR, LLP

Dated: 3-3-04

By: 

Ronald J. Schoenbaum
Registration No. 38,297
Attorney of Record
2040 Main Street
Irvine, CA 92614

M:\AMAZON\060A\AMEND1.DOC\dns
020404

AMAZON.060A PATENT

GRAMMAR GENERATION FOR VOICE-BASED SEARCHES

Background of the Invention

Field of the Invention

[0001] The present invention relates to ~~voice~~ speech recognition systems, and more particularly, relates to methods for recognizing utterances when a user performs a voice-based search.

Description of the Related Art

[0002] With the increasing popularity of wireless devices, many Web site operators and other content providers are deploying voice driven interfaces ("voice interfaces") for allowing users to browse their content. The voice interfaces commonly include "grammars" that define the valid utterances (terms, phrases, etc.) that can occur at a given state within a browsing session. The grammars are fed to a ~~voice~~ speech recognition system and are used to interpret the user's voice entry. In Web-based systems, the grammars are typically embedded as text files within voiceXML versions of Web pages. To support the use of multiple-term utterances, the grammar may include common phrases (ordered combinations of two or more terms).

[0003] One problem with ~~voice~~ speech recognition systems is that the reliability of the recognition process tends to be inversely proportional to the size of the grammar. This poses a significant problem to content providers wishing to place large databases of products or other items online in a voice-searchable form. For example, if all or even a significant portion of the possible word combinations are included in the grammar as phrases, the grammar would likely become far too large to provide reliable ~~voice~~ speech recognition. If, on the other hand, commonly used terms and/or phrases are omitted from the grammar, the system may be incapable of recognizing common voice queries. The present invention seeks to address this problem.

Summary of the Invention

[0004] The present invention provides a system and associated methods for generating a ~~voice~~ speech recognition grammar for interpreting voice queries of a database or other domain of items. The items may, for example, be book titles, movie titles, CD titles, songs, television shows, video games, toys, published articles, businesses, Web pages, users and/or any other type of object for which text-based searches are conducted. The invention is particularly well suited for conducting voice-based title searches. (As used herein, a "title search" is a field-restricted search in which items are located using terms appearing within item titles.)

[0005] In accordance with the invention, phrases are extracted from the searchable representations of the items, and are processed to identify (predict) the utterances that are most likely to occur within queries for such items. The phrases may, for example, include or be extracted from the titles of the items (e.g., to generate a grammar for interpreting voice-based title searches). Individual terms (e.g., single-term titles), may also be extracted from the items.

[0006] To identify the most likely utterances, each extracted phrase is exploded into its individual terms plus all forward combinations of such terms (i.e., ordered sets of two or more consecutive terms). For example, the phrase "the red house" would produce the following utterances: "the," red," "house," "the red," "red house," and "the red house." To avoid an undesirably large number of forward combinations, the extracted phrases may be limited in size to a certain number of terms. For example, if the grammar is derived from the titles of the items, a title having more than N terms (e.g., six terms) may be subdivided into two or more smaller phrases before phrase explosion.

[0007] A set of heuristics is applied to the resulting utterances to (a) remove utterances that are deemed unhelpful to the searching process (e.g., duplicate utterances, and utterances that would produce too many "hits"), and (b) to translate certain utterances into a format more suitable for use by the ~~voice~~ speech recognition system. The remaining single-term and multiple-term utterances are combined to form the ~~voice~~ speech recognition grammar. A relatively small set of "canned" utterances may also be inserted. The grammar thus contains single-term and multiple-term utterances derived from the items, with the

multiple-term utterances consisting essentially, or at least primarily, of forward combinations generated from the extracted phrases. The grammar is provided to a conventional ~~voice~~ speech recognition engine that is used to interpret voice queries for the items. The process of generating the grammar may be repeated as needed to maintain a grammar that is consistent with the contents of the database. Further, different grammars may be generated for different sets or domains of items.

[0008] An important aspect of the invention is that the resulting grammar tends to contain the terms and phrases most likely to be used within queries for the items, yet tends to be sufficiently small in size (even when the domain of items is large) to enable reliable ~~voice~~ speech recognition.

Brief Description of the Drawings

[0009] These and other features will now be described with reference to the drawings summarized below. These drawings and the associated description are provided to illustrate preferred embodiments of the invention, and not to limit the scope of the invention.

[0010] Figure 1 illustrates a process for generating grammars for use in voice based searches.

[0011] Figure 2 illustrates a process for performing a voice-based title search of a database.

[0012] Figure 3 illustrates a Web-based system in which the invention may be embodied.

[0013] Figure 4 illustrates an example implementation of the Figure 1 process for generating a title search grammar.

Detailed Description of a Preferred Embodiment

[0014] For purposes of illustrating one particular application for the invention, a particular embodiment will now be described in which ~~voice~~ speech recognition grammars are generated for interpreting voice-based searches, and preferably voice-based title searches, for products represented within a database (e.g., book, music, and/or video products). It will be recognized, however, that the invention may also be used for conducting searches for other

types of items, such as Web pages indexed by a crawler, downloadable software, companies, chat rooms, court opinions, telephone numbers, and other users. In addition, the invention may be used in the context of searches other than title searches, including but not limited to non-field-restricted searches, field-restricted searches corresponding to other database fields, and natural language searches.

[0015] In the context of the preferred embodiment, each item (product) is represented in the database as a record containing multiple fields, each of which contains a particular type of data (e.g., author, title, subject, description, etc.). The term "item" will be used generally to refer both to the products themselves and to the database records for such products. The term "title" will be used to refer generally to a name of a product such as the name of a book, CD, movie, song, article, toy, or electronics device.

[0016] Figure 1 illustrates the general process used in the preferred embodiment to generate a grammar for interpreting voice queries. The process may be executed as needed (e.g., once per week, when new items are added to the database, etc.) to ensure that the current grammar corresponds closely to the current contents of the database. Different grammars may be generated for different categories of items (e.g., books versus movies) to support category-specific searches. A more specific implementation of the Figure 1 process is discussed separately below with reference to Figure 4.

[0017] As depicted in Figure 1, the first step 102 of the process involves extracting or copying character strings from the text of the items in the searchable domain. These character strings necessarily include phrases (ordered combinations of two or more terms), and may include individual terms. Preferably, this step is performed by extracting the item titles in the searchable domain, and the resulting grammar is used to interpret voice-based title searches. For example, to generate a grammar for interpreting title searches for books, the title of each book in the database would be read. As will be recognized, the character strings could alternatively be extracted from other fields or portions of the searchable items to support other types of searches. For example, the strings could include or consist of one or more of the following: (1) complete sentences extracted from reviews of other descriptions of the items, (2) headings or sub-titles extracted from item text, (3) phrases determined by a text-processing algorithm to be characterizing of the respective items, and

(4) phrases which appear in bold or other highlighted item text. The extracted character strings are stored in a file or other data structure which, following completion of the process, represents the ~~voice~~ speech recognition grammar.

[0018] As depicted by steps 104 and 106, each character string (typically a phrase) is preferably preprocessed before the other process steps are performed. For example, phrases that exceed a predefined number of terms may be subdivided into smaller phrases (to avoid large numbers of forward combinations in step 108), and symbols and abbreviations may be converted to their word equivalents. An example set of pre-processing operations is discussed below with reference to Figure 4.

[0019] In step 108, each phrase is expanded or exploded into a set consisting of (a) all terms of the phrase (individually), and (b) all forward combinations, where a forward combination is defined as an ordered group of two or more consecutive terms of the phrase. For example, the phrase "the red house" would be expanded into the following set of character strings:

the
red
house
the red
red house
the red house

If the original character string had been divided into sub-phrases during preprocessing (step 106), the explosion step 108 is applied separately to each sub-phrase. The explosion step has no effect on any single-term character strings extracted in step 102.

[0020] After the explosion step 108 has been applied to all phrases, a preliminary version of the grammar exists. This preliminary grammar consists of both single-term and multi-term character strings, each of which represents an utterance that may be included in the final grammar. As depicted by step 110, a set of heuristics is applied to the entries within this preliminary grammar to (a) remove utterances that are deemed unhelpful to the searching

process (e.g., duplicate utterances, and utterances that would produce too many "hits"), and (b) to translate certain utterances into a format more suitable for use by the ~~voice~~ speech recognition system. A preferred set of heuristics for use in generating a title search grammar is described below with reference to Figure 4.

[0021] Finally, as illustrated by step 112, a small set of canned utterances may optionally be added to the grammar to handle special situations. The resulting set of utterances is then appropriately sorted (not shown), and is stored as a grammar file (step 114) for use by a ~~voice~~ speech recognition system.

[0022] An important characteristic of the resulting grammar is that the phrases contained within the grammar consist essentially, or at least primarily, of selected forward combinations of terms derived from the titles or other extracted phrases. Other combinations of terms are omitted from the grammar. For example, for the book title "Into Thin Air," the non-forward-combinations "air into," "thin into," "air thin," and "air thin into" would not be added to the grammar. Because users tend to utter only individual terms or forward combinations of terms when conducting voice searches, the process captures the terms and phrases that are most likely to be used. Further, because other phrases are generally omitted, the grammar tends to be sufficiently small to provide reliable speech recognition – even when the number of items in the searchable domain is large (e.g., hundreds of thousands or millions). Other processing methods that produce a grammar having such attributes are within the scope of the invention.

[0023] Figure 2 illustrates how the resulting grammar may be used to process a voice-based title search using voiceXML pages. This process may be implemented through executable code and associated content of a Web site or other system that provides voice searching capabilities. A conventional automated ~~voice recognition (AVR)~~ speech recognition (ASR) system that interprets voice according to externally supplied grammars may be used to implement the ~~voice~~ speech recognition tasks.

[0024] As depicted by Figure 2, after the user selects a title search option (step 202), the user is prompted (typically by voice, but optionally by text) to utter all or a portion of a title. For example, if the user is searching for the title "Disney's, the Hunchback of Notre Dame," the user could say "Hunchback of Notre Dame." The voice prompt, and the

corresponding grammar for interpreting the user's query, may be specified within a voiceXML page provided to the ~~AVR~~ ASR system using well-known methods.

[0025] As illustrated by steps 206 and 208, the ~~AVR~~ ASR system interprets the user's voice query by attempting to match it to an utterance contained within the grammar, and if a match is found translates the voice utterance into a corresponding text query. The text query may optionally include Boolean operators (e.g., the query terms may be explicitly ANDed together). In step 210 and 212, the text query is used by a conventional search engine to search a database of items, and the search results are returned to the ~~AVR~~ ASR system (for audible output to the user) within a voiceXML page.

[0026] Figure 3 illustrates a typical Web site system in which the invention may be embodied, and shows some of the components that may be added to the system to implement the invention. In this system, users can browse the Web site using either a conventional web browser (not shown) or using the site's voice interface. Users of the voice interface connect to the site by establishing a telephone connection to a conventional ~~AVR~~ ASR system 302 from a mobile or landline telephone 304 (or other device that supports the use of voice). The ~~AVR~~ ASR system 302 may, but need not, be local to the web server 306. Although the illustrated system uses voiceXML to provide the voice interface, it will be recognized that the invention is not so limited.

[0027] As depicted by Figure 3, the system includes an indexed database 307 of the works or other items for which searches may be conducted. A grammar generation processor 310 accesses this database 307 to generate grammars according to the process of Figure 1, a specific implementation of which is shown in Figure 4 for performing title searches. The grammar generation processor 310 is preferably implemented within software executed by a general-purpose computer system, but could alternatively be implemented within special hardware. Each grammar 316 is preferably stored as part of a title search page 314 that is passed to the ~~AVR~~ ASR system when a title search is initiated. As illustrated, a separate title search page 314 and grammar 316 may be provided for each category of items that is separately searchable (e.g., books, music, and videos). The title search pages 314 are stored within a repository of voiceXML content 312.

[0028] When a user submits a voice-based title search from a telephone 304 or other device, the ~~AVR~~ ASR system attempts to match the user's utterance to a textual utterance contained within the relevant grammar 316. If no match is found, the ~~AVR~~ ASR system may output an audible error message to the user, or may simply fail to respond. If a match is found, the ~~AVR~~ ASR system translates the voice query into an HTTP request specifying the user's query. The Web server 306 processes the request by invoking a query server 320 to search the database 307, and then returns to the ~~AVR~~ ASR system a voiceXML page specifying the search results. The search results may then be output in audible or other form to the user.

[0029] Figure 4 illustrates a particular implementation of the Figure 1 process as applied to item titles. The process consists of two stages: a per title stage 402 and a per grammar stage 404. In the illustrated embodiment, the process performs the per title steps 402 on each title while writing results to a grammar file, and then performs the per grammar steps 404 on the resulting grammar file. The output of the process is used to interpret voice-based title searches. As will be apparent, the specific rules applied within these steps may be varied to accommodate the particular search context (e.g., item category, number of items, etc.) for which the grammar is being generated.

A. Per Title Steps

[0030] The following is a description of the per-title steps.

[0031] In the pre-filtering step 412, symbols such as "&" and "+" are converted to their word equivalents ("and" and "plus" in this example). In addition, punctuation is removed from the title, and all terms are converted to lowercase. Further, predefined phrases that are deemed unlikely to be used within queries may be filtered out of the title.

[0032] Another pre-filtering operation that may be performed is to subdivide long titles into shorter phrases. For example, any title having more than six terms may be subdivided into phrases of no more than six terms. One method for dividing the titles involves using a language-processing algorithm to attempt to extract sub-phrases that each contain a noun, a verb, and an adjective. Each sub-phrase is thereafter processed via steps 414-425 as if it were a separate title.

[0033] In the numeric conversion step 414, Roman numerals and numeric phrases are converted into corresponding word representations. For example, the character string "21st" would be converted to the phrase "twenty first," and the string "XV" would be converted to "fifteen." Standard numbers preferably are not converted to word representations at this point, but rather are so converted following the explosion step 420. For instance the string "21" would not be converted to "twenty one" in this step 414.

[0034] In the abbreviation conversion step 416, abbreviations are expanded into their term equivalents. For instance, the abbreviation "Dr." is converted to "doctor," and "Mr." is converted to "mister."

[0035] In the duplicate phrase removal step 418, any duplicate phrases are removed from the title. For instance, in the title like "Game Boy, Plastic Case, Game Boy Color, for Game Boy II," two of the three "Game Boy" phrases would be removed.

[0036] In the explosion step 420, the process explodes any phrase extracted from the title into its individual terms and forward combinations, as described above. If the title had been divided into sub-phrases, each sub-phrase is exploded separately. If the pre-processed title consists of only a single term, no explosion processing is necessary. As discussed above, the explosion step has the effect of extracting the phrases that are most likely to be used in voice queries for the title, while omitting other word combinations that are less likely to be used (e.g., for the title "The Red House," the combinations "the house" and "house red"). The output of the explosion step 420 is a list one or more utterances that may be included in the grammar to enable the particular item to be located. As discussed below, some of these utterances may be removed or modified during subsequent steps of the process.

[0037] Following the explosion step 420, standard numbers are converted to their word counterparts (step 424). For example, the number "122" would be converted to "one hundred and twenty two." This conversion step is performed after the explosion step 420 so that the explosion does not produce an unnecessarily large number of utterances. For example, during the explosion step 420, the string "122" is treated as a single term, rather than the five terms appearing in "one hundred and twenty two."

[0038] In the acronym expansion step 424, acronyms are expanded into corresponding terms and phrases. For instance, "WWF" is converted to "W W F." Further, special cases like "4X4" and "3d" are converted to "four by four" and "three dee," respectively.

[0039] In step 425, the resulting list of utterances for the current title is added to a preliminary version of the grammar. The process is then repeated until all of the titles have been processed.

B. Per Grammar Steps

[0040] The following is a description of the per-grammar steps that are applied to the preliminary version of the grammar:

[0041] In the duplicate utterances removal step 428, duplicate utterances occurring within the grammar set are removed. For instance, if the titles "Disney's Hunchback of Notre Dame" and "Memoirs of Disney's Creator" both exist within the database, the utterance "disneys" will appear at least twice within the grammar. Only a single occurrence of the utterance is retained.

[0042] In the noise word removal step 430, specific grammar entries that are deemed non-useful to the search process (e.g., would produce a large number of search results) are removed from the grammar. For instance, in one embodiment, the following types of single-term utterances are removed: (a) colors, such as "red" and "green," (b) common words such as "is," "or," "for" and "like," and (c) numbers such as "five" and "twenty." A list of the noise words for a particular search domain can be generated automatically by identifying the terms that appear in more than a predefined threshold of titles (e.g., 20).

[0043] In the "special case removal" step 432, utterances that satisfy certain operator-defined heuristics are removed. For example, the following types of utterances may be removed: (a) utterances that end in words such as "and," "are," "but," "by," and "if," and (b) utterances containing nonsensical word patterns such as "the the," "the in," and "of for." A system operator may develop a set of heuristics that is suitable for a particular search engine and database by manual inspection of the grammars generated.

[0044] In the pre-canned grammar addition step 404, a predefined set of utterances is preferably added to the grammar. For instance, because the video game title "pokemon" is pronounced by some users as the three separate terms "pok," "ee," and "mon," it may be desirable to add these terms to the grammar as single-term utterances. Typically, only a relatively small number of utterances are added to grammar during this step.

[0045] The resulting grammar is stored as a text file or other data structure (step 436), and is provided to the ~~AVR~~ ASR system (e.g., within a voiceXML page) when a user initiates a voice-based title search.

[0046] As will be apparent, the steps of the above-described process can be varied in order without affecting the resulting grammar. For example, noise word utterances (step 430) and special case utterances (step 432) could be removed during or immediately following the explosion step 420.

[0047] One variation of the above-described process is to store within the grammar structure identifiers of the titles to which the utterances correspond. The grammar would thus serve both as a ~~voice~~ speech recognition grammar and as a search engine index. In such embodiments, the ~~AVR~~ ASR system 302 and the query server 320 may be combined into a single program module that uses the grammar/index structure to both interpret an utterance and look up the corresponding search results.

GRAMMAR GENERATION FOR VOICE-BASED SEARCHES

Abstract of the Disclosure

A grammar generation process generates a ~~voice~~ speech recognition grammar for interpreting search queries of a domain of items. The grammar comprises both single-term and multi-term utterances derived from the texts of the items (preferably the item titles). The utterances are derived in-part by expanding phrases selected from the item text into their individual terms plus all forward combinations of such terms. The forward combinations and individual terms that are deemed not useful to the search process are filtered out of the grammar. The process tends to produce a grammar containing the utterances that are most likely to occur within voice queries for the items, while maintaining a grammar size that is sufficiently small to provide reliable ~~voice~~ speech recognition.